

# Research and Analysis of Game Tactics Based on Data Mining Technology

ZHOU Li

Physical Education ,Jiangxi Science and Technology Normal University, Nanchang ,China

zhouljx@163.com

**Keywords:** Data Mining, Sports Competition, Tactical Analysis, Temporal and Spatial Relations

**Abstract.** Data mining is a technique that uses machine learning and statistical learning to discover potential rules from data. In view of the existing problems in sports competitions, the application principles of data mining technology in sports competitions are discussed, in this thesis work a system is designed which enables a coach to evaluate tactic by enabling him to express a tactical pattern and by visualizing the results. A prototype of the system is implemented to evaluate the design of the system. Data mining technology will guide the scientific development of sports competition, and will have broad application space.

## Introduction

The sports world is known for the vast amounts of statistics that are collected for each player, team, game, and season. There are also many types of statistics that are gathered for each – a basketball player will have data for points, rebounds, assists, steals, blocks, turnovers, etc for each game[1, 2]. This can result in information overload for those trying to derive meaning from the statistics. Hence, sports are ideal for data mining tools and techniques. Sports organizations, due to the extremely competitive environment in which they operate, need to seek any edge that will give them an advantage over others. It would appear that the culture has long encouraged analysis and discovery of new knowledge exhibited by its longstanding utilization of scouting. However, traditionally sports knowledge has been believed to be contained in the minds of its experts – the scouts, coaches, and managers. Only recently have sports organizations begun to realize that there is also a wealth of knowledge contained in their data. Currently, most team sports organizations employ in-house statisticians and analysts to retrieve meaning and insight for the scouts who evaluate future prospects and talent, the coaches who are in charge of the team on the playing surface, as well as the general managers who are in charge of drafting or signing players.

## Related work

Data mining can be used by sports organizations in the form of statistical analysis, pattern discovery, as well as outcome prediction[3, 4]. Since athletes are their biggest investments, teams are hoping that prediction of injury will help save millions of dollars. Similarly, researchers indicate that data mining can be used on physical aptitude test data in order to predict future physical performance[5]. Data mining software was used to link test data of cadets at the United States Military Academy and their actual performance in a required fitness class. This type of analysis would have significant implications to sports organizations – which put prospects through rigorous examination. Prior to each annual draft, the National Football League (NFL) holds an event referred to as the Combine in which eligible college players perform various performance tests in the presence of various team personnel including scouts, coaches, and general managers. Among the included physical tests are the 40 yard dash, vertical and broad jump, as well as physical measurements. Throughout the years, NFL teams and experts have developed common consensus on what are considered poor, good, and excellent results in the test based on the performance of the athletes throughout the previous years. In terms of the mental aspect of potential players, the Combine also allows teams to interview players and each player is expected to take the Wonderlic test. The fifty minute pre-employment examination is used by teams to assess the intelligence of

prospects. Similar to the physical tests, NFL teams have developed expected Wonderlic scores based on amount of intelligence presumed necessary to play particular positions [6].

### The basic technology of data mining

Data mining is a lot of, incomplete, noisy, fuzzy, random data is extracted from implicit in them, people do not know in advance, but is potentially useful in the process of knowledge. Data mining as an emerging multidisciplinary application, decision support are all walks of life activities play an increasingly important role. Data mining include clustering, classification, prediction, outlier analysis, association rules, and other aspects of visual description. Data mining association rule mining is an important research topic, but also one of the most widely used, in recent years has been the industry's research focus. The main purpose of association rule mining for sport tactics is to find valuable items between large data sets relevant contact, as shown in fig. 1.

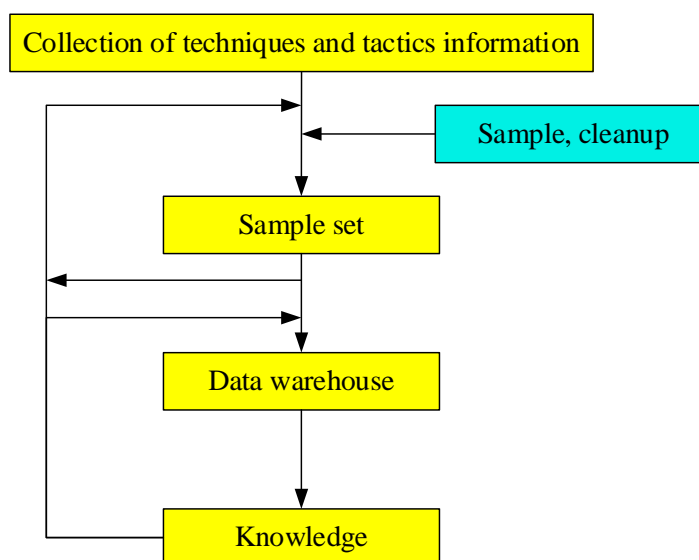


Fig. 1. The basic process of data mining

### Appliance of Data Mining in Sport

Sport is full of data. These data may show the individual qualities of a certain player, the events that happened during a game, and/or how a team is functioning as a whole. It is important to determine which data should be stored and a way to maximally use them<sup>[7]</sup>. By finding the right way to extract sense from these data and to transform them into practical knowledge, sports organizations provide themselves an advantage in comparison to other teams<sup>[8]</sup>. Different sports associations have different attitudes to data. This approach may be divided into five level<sup>[9]</sup>:

- There is no connection between sports data and their use
- The experts from a given field work on predictions using their instincts and hunches
- The experts from a given field work on predictions using data collected
- The use of statistics in the decision-making process
- The use of data mining in the decision-making process.

The first type of approach is when there is no connection between sports data and their use. These sport organizations often obtain certain data about players on their games and they ignore these. This is characteristic for amateur sports organizations, since their emphasis is on fun or on introducing the sports basics.

The next type of approach is based on an expert from a given field who is predicting based on his own experience. It used to be a widely accepted notion that these experts (coaches, managers, scouts) could efficiently use their insights and experience in order to reach the correct decisions. Decisions made from this type of approach are usually based on predictions or instincts, and not on

real data and information. These decisions may include taking certain types of actions or making certain player changes because such a decision "looks right".

The third type of approach is when experts start using collected data. Decisions at this level include playing certain players, for whom it has been proven that they cooperate well and are taking actions that score points more often.

The fourth type includes statistics in the decision-making process. These statistical measurements may be simple, such as the measurements of the frequency of certain events; or complex, which divides the performance of a whole team and assigning merits to each player in given game or a league. Statistics is used as a tool, helping experts in making correct decisions.

The fifth type uses data mining techniques. They do have potential to help predictions. Statistical techniques are still at the core of data mining, but statistics is used to separate a pattern or any other behavior of interest (the tendencies of opponent players) from the background noise. Statisticians do not explain relations within data, since this is a point of data mining. This type of approach has a potential to be used in order to help experts to make appropriate decision or to be used independently in order to make decisions without experts. Most sports organizations use the third or fourth type of approach between data and their use, and only a few use data mining techniques. Although data mining was introduced in sport relatively recently, the results of teams who apply these techniques are exceptional.

Although the use of statistics in the decision-making process is certainly a step up from decision-making based on instinct, statistics can also lead decisions in a wrong direction if there is no deep understanding of a problem. This tendency of statistics may be a consequence of imprecise performance measurements or of insistence by the sports community on certain characteristics. This we may support by the fact that a certain player may have exceptional individual statistics but may still have small influence on the team as a whole.

### **The establishment of domain model association data mining model**

**Temporal relations.** When evaluating the tactics of a game a coach is not interested in a specific action but in a combination of actions like the buildup towards scoring a goal, or how specific actions follow upon each other. In current match annotations, the actions are modeled as separate event entities which all have a timestamp. When a coach wants to know what the following action is he has to derive the following action by looking for the next action based on the timestamp. The information is available in the data but only implicit. In the domain model the data is modeled at a higher abstraction level and this temporal relation can be made explicit. Instead of looking at a single action an action has a predecessor and a successor. The predecessor is the action that happened before the action and the successor is the action that happens after the action. This is defined by the timestamp of the actions. Exceptions are the first action and the last action of a period. A match consists of two periods of 35 minutes. The first action of a period does not have a predecessor while the last action of a period does not have a successor. This enables reasoning about the temporal relations between the actions while hiding the timestamps.

**Spatial relations.** An action occurs at a certain location in the hockey field. The same as with the temporal relations the relations are made explicit in the domain model. This is done by deriving the spatial relations from the match annotation and mapping it to a class that describes the relations. The classification is performed based on the definitions of a coach and can be configured by a coach. This supports the flexibility requirements of the system. The process of the classification of the spatial relations is visualized in Fig.2.

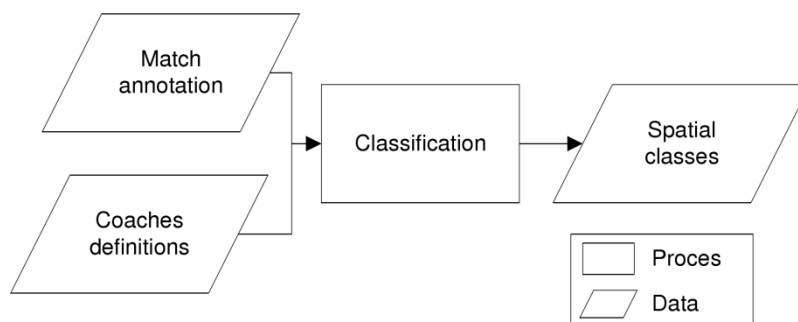


Fig. 2. Spatial relations classification process

**Domain model.** In Fig. 3 a class model visualizes the domain model. On the left side in the yellow square the subject properties are modeled. An action is performed by a player and a player has a position like defender or midfielder and a player is part of a team. The right blue square shows the derived spatial relations. The spatial derivate distance, direction and angle can only by derived from dynamic actions. The actions are linked together using a From and To relation, which makes the temporal relation between the actions explicit.

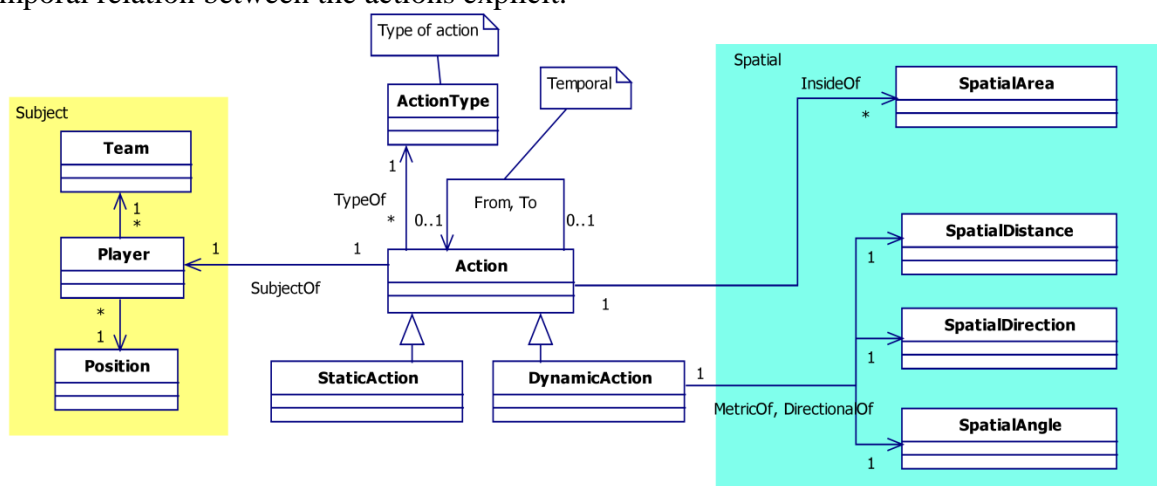


Fig. 3. Domain model

A task of the domain model is to express a tactical pattern. Via the domain study the components necessary to describe a tactical pattern are derived. These components are represented in the domain model. Therefore the domain model contributes to the expressiveness criteria of the system because with this domain model a tactical question can be expressed. In the domain model a level of abstraction is created. By making the spatial and temporal relations explicit as properties of an action the system offers a richer expression of an action. This enables the expression of domain concepts of field hockey like a “pass by player A to player B”. A downside to this abstraction is that the system has a lower expressive power because not every possible question involving the spatial and temporal information can be formulated anymore. Looking at flexibility criteria, the domain model offers the possibility for a coach to configure the spatial classification. This configurability enables a coach to create his own definitions for example the spatial lengths.

## Conclusion

This paper studies the difficulty of applying the data mining algorithm to the sports, and puts forward a scheme to analyze the technical and tactics of the sports game based on the data mining algorithm. The experimental results show that the scheme can be excavated by intuitive observation and simple statistics cannot be obtained, but for the coach Has important significance of the data, can tap the key action conversion process, the program for mining sports game tactical information have a certain role. The next step is to apply the algorithm to the analysis of the key regional transfer process, the key tactical transfer process, and the application of association rules and classification prediction in the sports game..

## References

- [1] Rani B., Kaur P., Bansal A. Analysis of student physical fitness data using data mining algorithm [J]. 2012, 7(11): 1659-1662.
- [2] Meir, Ronny. Appliarence of Neural Networks in Basketball Scouting [J]. Acta Polytechnica Hungarica, 2010, 7(4): 201-213.
- [3] Fielitz L., Scott D. Prediction of physical performance using data mining [J]. Research Quarterly for Exercise & Sport, 2003, 74(1): A25-A25.
- [4] Schumaker Robert P., Solieman Osama K., Chen Hsinchun. Sports Data Mining: The Field [M]. Springer US, 2010.
- [5] Fielitz Lynn, Scott David. Prediction of physical performance using data mining [J]. Research Quarterly for Exercise and Sport, 2003, 74(1): 1-25.
- [6] Schmittel Annelie, Sanderson Jimmy. Talking about Trayvon in 140 characters: Exploring NFL players' tweets about the George Zimmerman verdict [J]. Journal of Sport and Social Issues, 2015, 39(4): 332-345.
- [7] Acha Antonio Terroba, Kusters Walter A., Vis Jonathan K. Tactical Analysis Modeling through Data Mining - Pattern Discovery in Racket Sports; proceedings of the Kdir 2010 - Proceedings of the International Conference on Knowledge Discovery and Information Retrieval, Valencia, Spain, October, F, 2010 [C].
- [8] Li Hao Qin. Strategy and Analysis of Sport Events Based on Data Mining Technology [J]. Applied Mechanics & Materials, 2014, 687-691: 1137-1140.
- [9] Rey Ezequiel, Lagoballesteros Joaquín, Padróncabo Alexis. Timing and tactical analysis of player substitutions in the UEFA Champions League [J]. International Journal of Performance Analysis in Sport, 2015, 15(3): 840-850.